IJMRP

International Journal of Medical Research Professionals

Original Article

Comparison of Whole Genome of E.Coli O146 with Reference EPEC 2348/69

K Rajeshwari[1*], Beena Uppal[2], Rakesh Singh[3]

[1*]Professor, Department of Pediatrics, [2]Director Professor and Head, Department of Microbiology, Maulana Azad Medical College, New Delhi, INDIA.

[3]Principal Scientist, Division of Genomic Resources, NBPGR, New Delhi, INDIA.

Article History
Received: 09 Jan 2016
Revised: 11 Jan 2016
Accepted: 13 Jan 2016

***Correspondence to:**
Dr K.Rajeshwari,
Professor of Pediatrics,
Maulana Azad
Medical College,
New Delhi,India.
Rajeshwari.dr@gmail.com

ABSTRACT

Background: Escherichia coli are a common cause of acute diarrheal illness in children specially in developing countries. Enteropathogenic E.coli (EPEC) commonly affects infants and cause significant morbidity and mortality. An EPEC strain O146 was recovered from the stool of an infant admitted with acute diarrhea from a medical college hospital in New Delhi, India .Whole genome sequencing was done and the genome was compared to reference strain E.coli 2348/69.

Results: We sequenced the genome of Escherichia coli O146 using Illumina NextSeq 500 platform with 2 x 150 pair-end read chemistry. Approximately ~1 Gb of high quality reads of the sample were assembled using velvet software. We have obtained 188 contigs which comprising of ~4.9Mb genome. It has several genes that are related to virulence and pathogenicity that are responsible for causing disease .This was a typical EPEC having both eaea and bfpA gene. In addition this genome had many new genes responsible for antibiotic resistance which were not found in the reference genome EPEC 2348/69.

Conclusion: This E.coli O146 is a multiple drug resistant isolate recovered from a child with acute diarrhea .This is the first description of a comparative analysis of an EPEC isolated from a child with acute diarrhea with reference strain EPEC 2348/69.

KEYWORDS: E.coli O146, Reference based analysis of E.coli O146, Whole genome sequencing.

INTRODUCTION

In recent times advanced molecular techniques like Whole Genome Sequencing (WGS) has become available which is providing deeper insights into the pathogenesis, virulence genes, antibiotic resistance genes of enteropathogenic *Escherichia coli.* Whole genome sequencing is now available in many laboratories in advanced countries but in India it is still not widely available due to great costs involved.

Enteropathogenic E.coli (EPEC) has been implicated as the first pathotype of E. coli responsible for human disease especially pediatric diarrhea. Among the various strains E2348/69 belonging to O group 127 was studied using the method of whole genome sequencing. Throughout the world this has been labeled as a prototype to study virulence traits, physiology, pathology of diarrhea and for detailed molecular typing. Detailed analysis of this bacterium has led to the detection of Locus of Enterocyte Effacement (LEE) which codes for the Type 3 secretory systems.and various effector molecules which is responsible for adhesion and effacement.. lesions seen on the human gut epithelium which is pathognomonic of EPEC.

In a study by Iguchi et al. the whole genome sequence of E2348/69 was unraveled and it was compared with other pathotypes of *E.coli* discovered till now.[1] In this study, over 425 genes which were specific to E2348/69 were identified. It was deciphered that the genome of this strain is relatively less complex and has 20 type III secretion system effecting genes .The various virulence genes were identified and their association with various physiological processes were linked.

Using next generation sequencing methods, the DECA collection of *E.coli* was analyzed and draft genome sequences were published.[2] This contained information about 15 *E.coli* strains known to be commonly pathogenic to humans. These were enterotoxogenic

E.coli and are a very precious source of information for various microbiologists across the world and are used for various comparative genomic studies.

Studies of whole genome sequencing of EPEC isolated from pediatric diarrhea in India are nonexistent. Thus we studied an E.coli isolated from a child with acute diarrhea and compared it with EPEC 2348/69.

MATERIALS AND METHODS

Institutional ethical clearance was obtained from Maulana Azad Medical College Institutional Ethical Clearance committee.

The stool sample of the child (sample 535)was cultured using standard microbiological techniques and E.coli was isolated on Mc Conkey agar. For identification of EPEC, slide agglutination with antisera to common EPEC O antigens, was carried out. *E. coli* strain grown on a nutrient agar plate, was suspended in normal saline solution, autoclaved for 15 minutes and then examined by slide agglutination using commercially available antisera, in a kit identified as "Pathogenic *E. coli* Antisera" (Denka Seiken Co.,Ltd.,Tokyo,Japan). DNA isolation was done using MB505 HiPurA™ Bacterial and Yeast Genomic DNA Miniprep Purification Spin Kit procured from HiMedia Labs, Mumbai. This kit simplifies isolation of DNA from bacteria (Gram positive and Gram negative) and yeast by the spin column procedure. Next Whole Genome Sequencing was done and comparative analysis with EPEC 2348/69 was done as below.

Data generation

High quality data was generated using Illumina Technology. A total of ~1 Gb of data was generated and represented in table as below.

Table 1: Data Statistics

Description	Sample-535
Total number of Reads	3,114,429
Total number of data in Gb	~1.0 Gb

About Reference Genome

Escherichia coli O127:H6 str. E2348/69 Strains was used for reference based mapping as given below in Table 2.

Table 2 : Reference genome: strain and ID

Reference Genome	Number of bases (bp)	Reference sequence ID
***Escherichia coli* o127:H6 str. E2348/69**	4965553	NC_011601.1

Mapping Statistics

The high quality reads were mapped on different reference genome using CLC-Genomics Workbech6 with default parameters. After mapping, Samtools was used to call the consensus from the bam file generated from CLC-Genomics Workbench. The coverage and mapping reads percentage is given in Table3 and Table4 respectively.

Table 3: Mapping Statistics and Genome Coverage

Sample Name	Reference Genome	Genome Coverage (bp)	% Genome Coverage
Sample-535	*E.coli* str. E2348/69	3964476	80

Table 4: Statistics of Mapped reads

Mapping on reference genome	Total No. of reads	No of mapped reads	% of mapped reads
***E.coli* str. E2348/69**	3,114,429	2,459,264	78.96%

Genes Mapping

Table 5: Number of Genes in reference genome.

Description	No. of Genes
Escherichia coli O127:H6 str. E2348/69	4,886
Total no of genes present in sample with >60 coverage	3,834

A Total no of mapping reads percentage was highest on *Escherichia coli o127:H6 str. E2348/69*, its genes sequences were fetched. The high quality reads of the sample were mapped on gene sequences using CLC-Genomics workbench (version 6) with optimized mapping parameter. After mapping, SAMtool was used to call the consensus from the BAM file generated from CLC-Genomics workbench. The statistics are mentioned below and gene sequences are provided in supplementary files 1. The genes in the reference genome (*Escherichia coli O127:H6 str. E2348/69*) were aligned against our query sample, and the top hits rendered in a genome map using G-View software.

The two outer slots (brown is for query and red is for reference), It is showing a BLAST hit with query files. Empty regions on the query slots indicate areas where there were no BLAST hits between the reference and the query files. The next circle in black represents the GC content and purple color represents the GC skew in the consensus. The inner most circle represents the genes (Figure1)

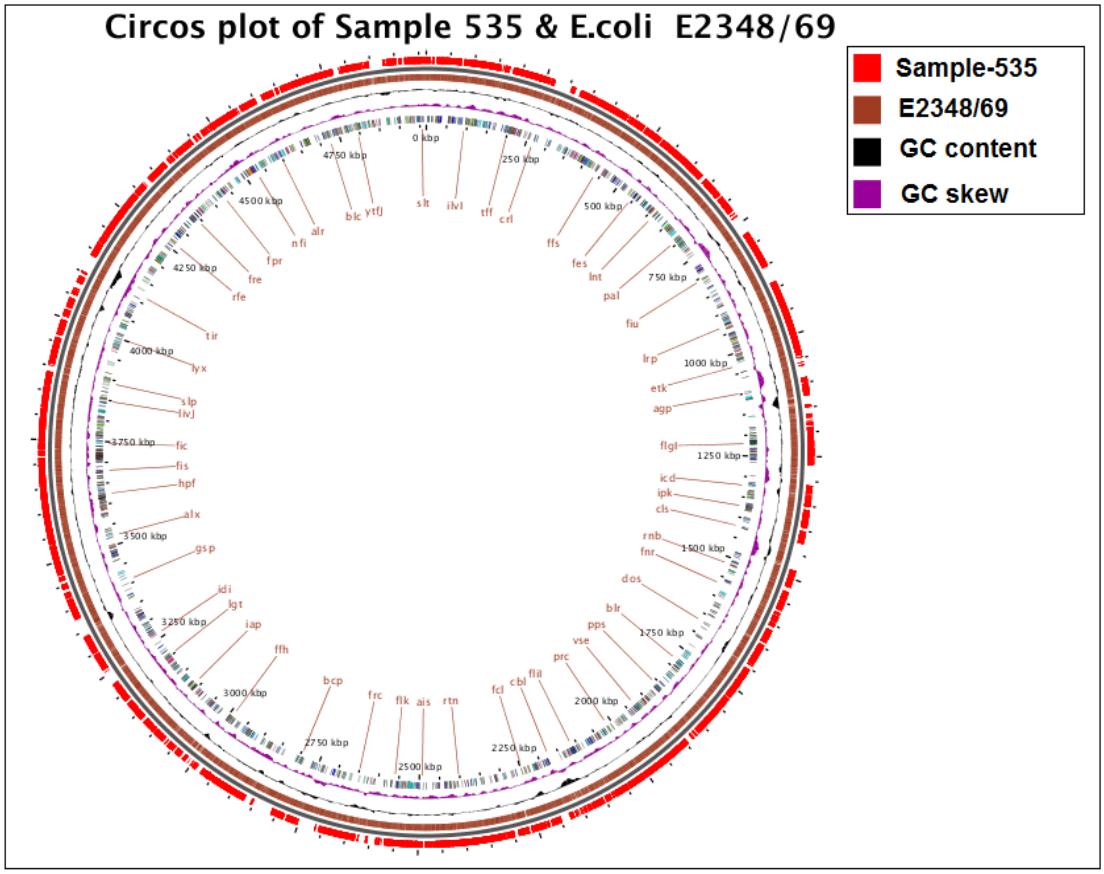

Figure 1: Circos Plot explaining the Genes in the reference and consensus sequence of sample-535.

Table 6: Showing Bacterial secretion system [PATH:ko03070]

CDS and contig number	Gene name	Protein
CDS_1545_contig_40	K03221 yscF	type III secretion protein SctF
CDS_1547_contig_40	K03222 yscJ	type III secretion protein SctJ
CDS_1537_contig_40	K03226 yscR	type III secretion protein SctR
CDS_1538_contig_40	K03227 yscS	type III secretion protein SctS
CDS_1539_contig_40	K03228 yscT	type III secretion protein SctT
CDS_1540_contig_40	K03229 yscU	type III secretion protein SctU
CDS_1541_contig_40	K03229 yscU	type III secretion protein SctU
CDS_1536_contig_40	K03225 yscQ	type III secretion protein SctQ

Summary

- Total high quality data of ~1 Gb for sample-535 was generated using Illumina technology.
- The mapping was performed for sample-535 against E.coli 2348/69 genomes as mentioned in Table4 CLCGenomics workbench. This resulted in 80% genome coverage
- Genes were identified for sample-535 using E.coli 2348/69 genes. A total of 3,834 genes were identified having greater than 60% coverage.

Supplementary Data files

- **Sample_535_gene_Coverage.xlsx:** This file contains gene of reference genome which were covered 60% or more by mapping reads generated for 535 sample.

DISCUSSION

This strain has various genes as shown below which codes for type III secretion proteins.The Type III secretory system are very complicated structures in bacteria that confer virulence characteristics in gram negative pathogens. This is done by injecting effector bacterial proteins into the cytoplasm of the host cell .This in turn induces changes in cytoskeleton of the host cell and enables attachment and blockade of various processes in the cytoplasm, causes cytotoxicity and alterations in the immune system. The pathologic hallmark of EPEC intestinal pathology is the attachment and effacing (A/E) lesion, whose formation is critically dependent on a T3SS coded within the locus of enterocyte effacement (LEE).

The various genes responsible for type III secretion proteins identified in the sample are as in Table 6.

Agglutination using O specific antisera confirmed O146 serotype and was identified as EPEC by the presence of the LEE (locus of enterocyte effacement) and the bfp gene. Attaching and effacing Escherichia coli are characterized by the presence of the locus of enterocyte effacement (LEE) pathogenicity island[3] and express the bundle forming pilus gene (bfp).[4] EPEC also has various secreted effector molecules by bundle-forming pili (BFP), EspC and eae genes.[5,6] This strain also has genes involved in tight association of these pathogens with the host cell, virulence responsible genes, and those conferring multidrug resistance.[7] Also present are genes responsible for different secretion mechanisms (type I, II and III) and causing transfer of secretory products across the inner cytosolic and outer membranes. There is in addition presence of multidrug transporter (mdtABCD) gene responsible for drug resistance.[8] This also has the astA gene, an additional gene causing virulence. This is the first report of complete genome sequence of *E. coli* O146 characteristically expressing a virulent genes and multidrug resistance genes.

CONCLUSIONS

a) This E coli strain had a adhesn gene which is seen in enteropathogenic E coli.

b) It also had a type III secretion apparatus gene coding for type III secretion apparatus protein.

c) E. coli heat-stable enterotoxin1 (EAST1) encoded by astA gene has been found in this enteropathogenic E. coli strain.

To conclude this is a atypical EPEC strain expressing additional virulence factors not encoded by LEE region. This is similar to a previous study by Silva et al ,where EAST1 was the most frequent (24%) virulence factor found in a collection of 65 aEPEC strains, and was significantly associated with childhood diarrhea.[9] Also O146 has not been studied so far in any part of the world. This is the first report of whole genome sequence of E .Coli O146 causing infantile diarrhea and expressing a toxigenic gene and multidrug resistance.

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession JWHN00000000. The genome is available for complete viewing at NCBI website.

CONFLICT OF INTEREST: None declared.

FUNDING: This is a non-funded genomic research.

REFERENCES

1. Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR et al. Complete Genome Sequence and Comparative Genome Analysis of Enteropathogenic *Escherichia coli* O127:H6 Strain E2348/69. J Bacteriol 2009;191(1):347:354.

2. Hazen TH, Sahl JW, Redman JC, Morris CR, Daugherty SC, Chibucos MC et al. Draft genome sequences of the diarrheagenic Escherichia coli collection. J Bacteriol 2012; 194(11):3026-7.

3. Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ,Black RE et al. Global Causes of Diarrheal Disease Mortality in Children <5 Years of Age: A Systematic Review. PLoS ONE 2013; 8(9): e72788. doi:10.1371/journal.pone.0072788.

4. Langendorf C, Le Hello S, Moumouni A, Gouali M, Mamaty AA, Grais RFet al.Enteric bacterial pathogens in children with diarrhea in Niger: diversity and antimicrobial resistance. PLoS One. 2015 Mar 23; 10(3):e0120275. doi: 10.1371/journal.pone.0120275. eCollection 2015.

5. Dutta S, Guin S, Ghosh S, Pazhani GP, Rajendran K,Bhattacharya MK et al. Trends in the Prevalence of Diarrheagenic *Escherichia coli* among Hospitalized Diarrheal Patients in Kolkata, India. PLoS ONE 2013;8(2): e56068. doi:10.1371/journal.pone.0056068.

6. Casalino M, Yusuf MW, Nicoletti M, Bazzicalupo P, Coppo A,Colonna B et al. A two-year study of enteric infections associated with diarrheal diseases in children in urban Somalia. Trans R Soc Trop Med Hyg 1988;82: 637–641.

7. Echeverria P, Orskov F, Orskov I, Knutton S, Scheutz F, Brown JE. Attaching and effacing enteropathogenic *Escherichia coli* as a cause of infantile diarrhea in Bangkok. J Infect Dis 1991; 164: 550–554.

8. Chen HD, Frankel G . Enteropathogenic *Escherichia coli*: unravelling pathogenesis. FEMS Microbiol Rev2005; 29: 83–98.

9. Silva LE, Souza TB, Silva NP, Scaletsky IC. Detection and genetic analysis of the enteroaggregative *Escherichia coli* heat-stable enterotoxin (EAST1) gene in clinical isolates of enteropathogenic *Escherichia coli* (EPEC) strains. *BMC Microbiology*. 2014;14:135. doi:10.1186/1471-2180-14-135.

Copyright: © the author(s) and publisher IJMRP. This is an open access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cite this article as: K Rajeshwari, Beena Uppal, Rakesh Singh. Comparison of Whole Genome of E.Coli O146 with Reference EPEC 2348/69. Int J Med Res Prof. 2016, 2(1); 70-73.